IN THE UNITED STATES DISTRICT COURT

FOR THE DISTRICT OF OREGON

UNITED STATES OF AMERICA,

v.

ODELL TONY ADAMS,

Defendant.

Case No. 3:19-cr-00009-MO-1

OPINION AND ORDER

MOSMAN, J.,

Defendant, Odell Tony Adams, filed this Motion in Limine [ECF 55] to consider the admissibility of scientific expert testimony relating to toolmark comparison evidence under *Daubert v. Merrell Dow Pharm., Inc.* 509 U.S. 579 (1993). As explained below, I GRANT Defendant's motion in part and exclude most of the disputed evidence.

FACTUAL BACKGROUND

Mr. Adams is charged with one count of Felon in Possession of a Firearm, in violation of 18 U.S.C. § 922(g)(1). Second Superseding Indictment [ECF 65] at 1. The charge stems from two distinct acts of possession of two different firearms, which Mr. Adams allegedly possessed on different days in different locations. *Id.* at 1-2; Compl. [ECF 1] at 3-4. One of the firearms is a Taurus model PD 24/7 Pro .40 caliber handgun. Indictment [65] at 2. The other is a Ruger model SR-40 .40 caliber handgun. *Id.*

The Government alleges that Mr. Adams shot at someone at the Speakeasy Lounge in east Portland on the night of October 5, 2018, after which he fled the scene. Gov't. Trial Br.

[ECF 82] at 2-9. He was not caught on the night of the shooting, but officers and investigators from the Portland Police Bureau (PPB) responded to the scene. *Id.* at 9; Compl. [ECF 1] ¶ 5. They took witness statements and seized a number of shell casings,[1] which corresponded to the number of bullet strikes found on an SUV and a Buick in the Speakeasy parking lot. Trial Br. [82] at 9. PPB officers then used surveillance videos from the Speakeasy to identify the individuals present at the scene, including Mr. Adams himself. *Id.* They then obtained his address and cell phone number from his federal probation officer. *Id.*

On October 31, 2018, PPB obtained a search warrant for Mr. Adams's residence, which authorized the seizure of Mr. Adams for his DNA and any evidence related to the Speakeasy shooting. Compl. [1] ¶ 5. Officers executed that search warrant on November 2, 2018. *Id.* ¶ 6. They arrested Mr. Adams and swabbed his DNA. *Id.* His roommates showed the officers to Mr. Adams's room and asserted that none of them owned or possessed any firearms. *Id.* In the bedroom, officers found clothing matching the description of that worn by the shooter, and they opened a crawlspace access panel located on the wall of the bedroom. *Id.* ¶¶ 7-8. Inside, they found the Taurus and Ruger handguns. *Id.* ¶ 8.

Investigators submitted the guns for forensic comparison to the shell casings found at the scene, with somewhat inconsistent results. The initial comparison test done by the National Integrated Ballistic Information Network (NIBIN) showed that, although the casings at the Speakeasy were .40 caliber, they were not shot from either the Taurus or the Ruger found in Mr. Adams's bedroom. Compl. [1] ¶ 9; Def.'s Mot. Ex. 3 [ECF 55-3] ("Police Report") at 1. PPB

---

[1] The witnesses and parties have described what's left over after the bullet is expelled as "shells," "shell casings," "casings," and "cartridges." The words are used somewhat synonymously, although technically a cartridge is the entire package of shell, primer, powder, and bullet. But in this opinion, I adopt the parties' practices of using all of these terms to mean "shell."

resubmitted the casings for testing by the NIBIN after discovering a possible error during the initial test. Police Report [55-3] at 7. The second test concluded that the casings were a "presumptive match" to the Taurus. *Id.* The PPB then submitted the casings for further forensic comparison testing at the Oregon State Police Crime Lab. *Id.*

Forensic scientist Travis D. Gover conducted the comparison testing for the OSP lab. Expert Witness Summ. [ECF 53] at 2. Mr. Gover concluded that the Taurus was an operable firearm and that the shell casings at the scene had been fired by the Taurus. Gov't. Ex. A-1 [ECF 53-2] at 1. In reaching that conclusion, Mr. Gover used what is known as the "AFTE methodology," which is used by members of the Association of Firearm and Toolmark Experts. Tr. [ECF 70] at 20. This methodology takes place in several steps.

The first step is to compare the class characteristics of the shell casings found at the crime scene and the firearm being tested, including the caliber of both, the general types of marks on the casings (linear or circular), any other features determined by the manufacturer, and the operability of the firearm. *Id.* at 20-21. The first step typically allows the examiner, by informed observation, to determine the manufacturer, the model, and the caliber of the firearm that expelled the shell casings. The second step is to test fire the gun being examined. *Id.* at 21. Mr. Gover said he does this by firing the gun into a water tank and collecting the bullets and casings expelled by the firearm. *Id.*

The second step allows the examiner to give his opinion as to the firearm's operability and provides the exemplars for comparison – *i.e.* the bullets and shell casings that can be compared to the suspected matches. Mr. Adams did not object to the methods or the conclusions involved in the first two steps.

The third step is to conduct the forensic comparison between the test-fired shells and the crime scene shells. *Id.* Mr. Gover said this is done by using a three-dimensional comparison microscope, which has two stages[2] connected to an optical bridge that allows the view to see each cartridge case side-by-side in one field of vision, like a split screen. *Id.* The images can then be magnified to varying degrees. *Id.* Mr. Gover said he first examines the test-fired shell casings for unique striations on the rim, where the hammer strikes the primer. Striations are marks that are specific to that firearm that would be reproduced on any shell casing fired from that gun. *Id.* at 21-22. He then compares the test-fired shell casings to the casings from the crime scene in order to see if they have striations or other characteristics that correspond to each other. *Id.* at 22.

If the examiner finds what is called "sufficient agreement," then he concludes there is a match—i.e., the shells from the crime scene were fired from the gun used in the test fire. The concept of sufficient agreement turned out to be the central problem with Mr. Gover's testimony. It starts with a comparator; something called the "best known non-match." In order to determine whether the degree of correspondence between the test-fire shells and the crime scene shells amounts to sufficient agreement, the starting point is that it must exceed the degree of correspondence of the best known non-match. The best known non-match, in a case involving shells, would be shells fired from two different firearms of the same manufacturer, model, and caliber made at the same plant on the same day, in immediate sequence—one right after the other. The underlying concept, as Mr. Gover explained it, is that this would produce two shells from two different firearms with the least possible degree of variation. *Id.* at 23-24. The

---

[2] The stage of a microscope is the platform where a specimen is placed for observation, often mounted to a glass slide.

examiner finds sufficient agreement, and therefore a match, if the test fired shells and the crime scene shells have greater correspondence than the best known non-match.

Mr. Gover used this methodology to determine that the shell casings from the Speakeasy shooting matched the Taurus found in Mr. Adams's bedroom wall, and the government offered his testimony and his report as evidence. Gov't. Expert Witness Summ. [ECF 53]; Gov't. Ex. A-1 [ECF 53-2] at 1.

PROCEDURAL BACKGROUND

As described above, the Government charged Mr. Adams with one count of Felon in Possession of a Firearm, alleging that Mr. Adams possessed two weapons of different makes—the Taurus and the Ruger handguns found in Mr. Adams's bedroom wall. Indictment [65] at 1-2; 18 U.S.C. § 922(g)(1). In preparation for trial, the Government intended to rely on Mr. Gover and his testimony to prove that Mr. Adams had possessed the Taurus firearm. Gov't. Expert Witness Summ. [ECF 53] at 2-3.

Mr. Adams filed a Motion *in Limine* in response to the Government's proffer of Mr. Gover as an expert witness, seeking a hearing pursuant to Federal Rule of Evidence 702[3] and *Daubert* to determine the admissibility of Mr. Gover's expert testimony and his ultimate

---

[3] Federal Rule of Evidence 702 provides:

> "A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:
>
> (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
>
> (b) the testimony is based on sufficient facts or data;
>
> (c) the testimony is the product of reliable principles and methods; and
>
> (d) the expert has reliably applied the principles and methods to the facts of the case."

conclusion that the shell casings matched the Taurus. Def.'s Mot. [55] at 1-2; *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 593-94 (1993). Mr. Adams argued that the kind of test Mr. Gover conducted is inherently subjective[4] and has been shown to lack reliability, which should disqualify it as expert scientific testimony. Def.'s Mot. [55] at 11. He requested that Mr. Gover's testimony be either limited in scope or excluded entirely. *Id.* at 2.

This court held a total of four hearings pursuant to Federal Rule of Evidence 702 and *Daubert*. Min. of Proceedings [ECF 61, 64, 75, 112]. At the first hearing, on December 4, 2019, Mr. Gover attempted to explain to the court how he arrived at his conclusion that the shell casings at the scene of the Speakeasy shooting were a match to the Taurus handgun found at Mr. Adams's residence. Tr. [ECF 70] at 12-66. In explaining his methodology, Mr. Gover struggled to articulate an objective basis for his conclusion or any means of measurement. In fact, he and the Government both referred to his methodology as "subjective" multiple times when pressed for a concrete explanation. For example:

> [PROSECUTOR]: And is it fair to say that AFTE [Association of Firearm and Toolmark Experts] has described the sufficient agreement as being a subjective standard?
>
> MR. GOVER: Yes. The interpretation of the objective observations, which is viewing those contours, surface contours of the toolmarks, being able to line up the peaks and valleys and ridges and furrows, that's—those are all objective

---

[4] In the testimony in this case, and in discussions in other published opinions, the *Daubert* issue is sometimes addressed as though the critical distinction is between admissible objective opinions and inadmissible subjective ones. That's not quite right. There are acceptable scientific opinions based on some limited degree of subjectivity; for example, a subjective decision about which underlying assumptions are better than others. The problem is when the subjective element of the methodology is also inscrutable—that is, when there is an inadequate explanation for the subjective component.

In this case, both the subjectivity and the inscrutability pose serious problems under *Daubert*. And I understand Defendant's challenge to have raised both issues, sometimes collapsed into the single word "subjective."

observations. The interpretation of the observation is what is the subjective aspect of the firearms identification.

Tr. [70] at 26.

. . .

[PROSECUTOR]: Yeah. Why isn't there a numeric – why can't we qualify this with numbers?

MR. GOVER: Part of it for me, to understand that quantifiable is almost another subjective aspect of when I'm looking at these furrows and ridges, what constitutes the consecutively—what they refer to as the consecutively manufactured or consecutively matching striae. So I don't have a complete answer to that and I've tried to formulate an argument as to why or why not, and I haven't, you know, heard a lot of other people talk about it.

Id. at 40.

. . .

THE COURT: What's the—what is the standard that an examiner uses to decide if there's a match or not?

[PROSECUTOR]: I think they're applying the AFTE methodology, and based on—it's a subjective standard, the government concedes that, but given the other indicia of reliability, the government's position is that—the other factors in the *Daubert* test should overcome the fact that it's a subjective standard as to that.

Id. at 67.

As the Government conceded, Mr. Gover could not articulate an objective means by which to determine whether the shells matched the firearm. The defense emphasized this subjectivity as the "most compelling" basis for its objection to Mr. Gover's testimony. *Id.* at 74. I declined to make a ruling at that hearing, but I did suggest to the parties that there may be some limitations put on Mr. Gover's testimony. *Id.* at 75-76.

At the second hearing, on December 12, 2019, Mr. Gover took the stand in order to amend his testimony from the December 4 hearing. Tr. [ECF 71] at 5. He explained that one component of firearms toolmark testing focuses on "consecutively matching striae" (CMS) and does require an objective measurement. *Id.* at 6-7. Specifically, the forensic examiner must find at least two sets of three matching striae in order to conclude the shell casing matches the firearm

in question. *Id.* at 7. However, Mr. Gover also testified that he did not actually quantify the matching striae in this case. *Id.* at 10-11. Rather, he relied on his experience to, essentially, eyeball the marks on the shell casings in order to determine whether they appeared to be sufficiently similar to constitute a match. *Id.* He could not say that he had made any quantitative findings. *Id.*

The Government agreed that Mr. Gover's ultimate opinion was subjective and that only this one component of the methodology was objective and quantifiable. *Id.* at 24. But this CMS measurement presented two problems: (1) it had not been shown to be independently reliable as a methodology, and (2) Mr. Gover did not actually use it in this case. *Id.* at 24-25. In fact, he had testified at the first hearing that the CMS method was "not universally accepted and used by the entire scientific community at this point" and that it could not be used as a stand-alone means of identification. Tr. [70] at 25. He then testified that he does not use that method routinely. *Id.*

In light of these inconsistencies in Mr. Gover's testimony, I ordered that we hold a third hearing to analyze the CMS method under *Daubert*. Tr. [71] at 27-28. I made it clear that the CMS method was now the only methodology I would consider allowing Mr. Gover to present to the jury as a possible match. *Id.*

We scheduled the third hearing for January 13, 2020. Min. of Proceedings [ECF 75]. On January 9, 2020, the Government notified the court it intended to withdraw Mr. Gover as an expert witness. Expert Notice [ECF 72]. The Government stated that it no longer intended to introduce any opinion from Mr. Gover that the shell casings matched the Taurus and that it would introduce only fact testimony from Mr. Gover. *Id.* at 2. However, the Government also argued that Mr. Gover could still be called as an expert witness under FRE 702, in order to explain the "procedures involved in the collection, examination, documentation, and analysis of

the firearms and shell casings in this case." *Id.* at 2-3. Finally, the Government argued that this withdrawal notice rendered Mr. Adams's original *Daubert* motion moot. *Id.* at 8.

At the January 13 hearing the parties presented arguments regarding the scope of Mr. Gover's testimony. Tr. [ECF 76] at 5-14. The Government argued that Mr. Gover should be permitted to testify to his methodology and that the photographs of the shell casings should be presented to the jury without Mr. Gover stating a conclusion about whether the photographs showed a match. *Id.* at 5-13. Mr. Adams argued that merely withdrawing Mr. Gover's conclusion—without withdrawing the remainder of his purportedly scientific testimony—did not moot the *Daubert* motion. *Id.* at 13-14. Further, he argued that presenting this evidence to the jury without any guidance about what conclusion to draw from it would be subjective and confusing to the point of being unhelpful. *Id.* at 14.

I ruled that this proffer from the Government did not fully moot the *Daubert* issue. *Id.* at 14-16. I then issued a ruling on Mr. Adams's motion in which I permitted the Government to introduce evidence of the make and model of the gun that fired the shell casings, the circumstances of their discovery, a general statement that Mr. Gover did a thorough examination of the shells, an opinion that the Taurus could not be excluded as the gun that expelled the shell casings, and that the Government could show the photographs as a demonstrative aid to the jury. *Id.* at 22-23. I then issued an order granting Mr. Adams's motion in part and denying in part, in light of the Government's withdrawal and in light of the testimony I did allow from Mr. Gover. Min. of Proceedings [75].

At the pretrial conference on February 10, 2020, I revisited the element of my ruling that allowed Mr. Gover to testify that he could not exclude the Taurus as the weapon at the scene of the shooting. Tr. [ECF 120] at 24-39. I asked the Government to confer with Mr. Gover and

determine whether he could provide a basis for that opinion that did not rest on the same methodology as his original opinion. *Id.* at 36. The Government responded in an email that stated that it intended to withdraw its proffer of any opinion evidence from Mr. Gover and offer only the following observational evidence:

- TAURUS Pistol: 40 caliber, semi-automatic pistol with a hemispheric-tipped firing pin, barrel with six lands/grooves and right twist;
- Casings Test fired from the TAURUS: 40 caliber, hemispheric firing pin impressions;
- Casings seized from outside the shooting scene: 40 caliber, hemispheric firing pin impressions;
- Bullet recovered from gold Oldsmobile: 40/10mm caliber with six lands/groves and right twist.

Email from Paul Maloney, 2/11/2020. Mr. Adams filed a motion to exclude this evidence. Mot. to Exclude [ECF 115]. I held a hearing on that motion, in which I concluded that this new, limited set of conclusions by Mr. Gover were based on direct observations and did not suffer from the inscrutability of his earlier conclusions. After some discussion about whether this was new testimony proffered too close to the start of trial, I held that the government could offer this limited observational evidence. This constituted my ultimate ruling on Mr. Gover's testimony.[5]

DISCUSSION

I. The *Daubert* Standard

The court performs a gatekeeping role by assessing the reliability of any expert testimony proffered under FRE 702. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999). FRE 702 and

---

[5] The Government asserted at the hearing that Mr. Adams had received the bullet during discovery. But the Government asserted in its trial brief that "[a]lthough some bullet fragments were recovered [at the Speakeasy], none were suitable for forensic analysis." Gov't. Trial Br. [ECF 82] at 9. I reserve the right to further modify my ruling, if necessary, in light of this discrepancy.

subsequent caselaw recognizes two general categories of experts: scientific experts and experts offering "technical" or "other specialized knowledge." *Kumho Tire*, 526 U.S. at 151; *United States v. Hankey*, 203 F.3d 1160, 1168 (9th Cir. 2000); *United States v. Alatorre*, 222 F.3d 1068, 1101 (9th Cir. 2000) (construing *Kumho Tire* and *Hankey*). *Daubert* lays out the test for whether scientific testimony is reliable. *See Alatorre*, 222 F.3d at 1100. In order to qualify as science, a proposition "must be derived by the scientific method." *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 593-94 (1993). The expert's assertion need not be verified as an objective certainty, but it must have been derived by a verified scientific process in order to meet the necessary standard of evidentiary reliability. *Id.* When assessing the reliability of scientific expert testimony, federal courts apply a five-part balancing test. *Daubert* 509 U.S. at 593-94. The five factors are: (1) whether a scientific theory has been tested; (2) whether the theory has been subject to peer review and publication; (3) the known or potential rate of error; (4) the existence and maintenance of standards controlling the technique's operation; and (5) general acceptance within the relevant scientific community. *Id.* In contrast, courts are not required to apply these factors when evaluating non-scientific expert testimony, although they may do so where the factors are applicable. *Kumho Tire*, 526 U.S. at 150; *U.S. v. Hankey*, 203 F.3d 1160, 1168 (9th Cir. 2000) (construing *Kumho Tire* to allow greater flexibility when determining the reliability of non-scientific expert witnesses).

Here, the Government offered Mr. Gover as a scientific witness and explicitly stated on the record that it did not intend to offer him as a technical witness under *Kumho Tire*. Tr. [70] at 53. Mr. Gover's title is "forensic scientist" and he works for the OSP Forensic Lab. Gover Curriculum Vitae [ECF 53-1] at 1. It is clear from this context that the proper analysis of his testimony occurs under *Daubert*, as the jury is likely to give special weight to his testimony

based on the apparent weight of scientific authority behind it. And, in fact, this is exactly what the Government wants from Mr. Gover's testimony. The Government argued that it needed to present Mr. Gover as an expert scientific witness to combat the "CSI effect," or the jury's expectation that solving crimes involves conclusive, scientific, forensic testing. Tr. [ECF 76] at 19. Without this kind of evidence, the Government argued, it would be at a severe disadvantage with the jury. *Id.* In light of this context—including the Government's own strategy for its case in chief—I must apply the *Daubert* factors and examine whether the proposed testimony gets past the gate of Rule 702 as being "scientific."

A brief introductory summary of the *Daubert* factors is helpful. The first factor, testability, refers to whether a methodology is falsifiable and replicable. *Daubert*, 509 U.S. at 593. This is what "distinguishes science from other fields of human inquiry." *Id.* (internal quotations omitted). Second, the peer review factor, asks whether a methodology or hypothesis has been published to the scientific community for the purpose of detecting substantive flaws. *Id.* This factor is relevant but not dispositive. *Id.* at 594. Third and fourth, the error rate and the standards controlling the technique's operation, are also not dispositive but are probative of whether a method is reliable. *Id.* Finally, "general acceptance" in the scientific community does not refer to any particular rate of acceptance, but a technique that has attracted only minimal support in the scientific community "may properly be viewed with skepticism." *Id.* (citing *United States v. Downing*, 752 F.2d 1224, 1238 (3rd Cir. 1985) (citation omitted)).

II. *Daubert* as applied to Mr. Gover's testimony

As the factual description of Mr. Gover's testimony shows, it is possible, at a superficial level, to conclude that his methodology satisfies the *Daubert* requirements. Replicability? Check.

Other forensic examiners trained in ballistics comparisons can perform an examination using the same basic methodology. In fact, double check: a forensic examiner in this very case did a second comparison and came to the same conclusion. Error rate? Check. Testing seems to indicate an error rate hovering around two percent. Testing and standards? Check, at least as to some of what is done. Publication? Check. Journals put out by associations of forensic examiners have published hundreds of relevant articles. Acceptance? Check. All reputable forensic examiners accept the ballistics comparison method used in this case as valid.

It is only when you look beneath the surface that the problems with Mr. Gover's methodology begin to emerge. It is worth remembering, at this point, that what we are analyzing here is the admissibility of *scientific* evidence. It is certainly possible to know things—and to know them so well that you are almost never wrong—without resorting to scientific methods. Consider the case of the art examiners in Malcolm Gladwell's book "Blink." Gladwell describes a statue that we eventually learn is a forgery, but which has passed several objective tests, including examination under an electron microscope, electron microprobe, mass spectrometry, X-ray diffraction, X-ray fluorescence, and mineral dating. Malcolm Gladwell, Blink 4 (2007). But after the scientists concluded that the statue was a genuine article of ancient Greek sculpture, several highly experienced art experts still sensed, for reasons they could not articulate, that it was a fake. *Id.* at 5-6. And they turned out to be right. *Id.* at 7-8.

Imagine, for a moment, a school to train such art experts. It is given the name, "The Academy for the Science of Art Examination." Everyone there is trained to follow a particular protocol, with several objective elements of scientific examination: carbon dating, microscopic examination, x-rays, and the like. The Academy establishes a journal, in which its graduates regularly publish papers describing their work in peer reviewed articles. There is regular testing

in which examiners are sent two works of art and told to determine which one is a forgery, a test on which the examiners have only a two percent rate of false positives—*i.e.* of misidentifying original works as forgeries. And they all agree that what they are doing is science. The only rub is this: when all the objective work is done, and it comes time to decide whether what they are looking at is fake, they are told to call upon all the experience they have, and the feeling that true art invokes in them, and then make a call by listening to their inner voice telling them what is right. The ultimate support for their conclusions, then, is almost entirely subjective and inscrutable.

Like any analogy, it does not map perfectly over our facts. But the central point is this: If you have elements of objective examination, and a lot of training, and you form a like-minded group to review each other's work and publish papers, and if you maintain standards to guide the objective elements of the process, and ensure that a certain protocol is followed, and if you are frequently correct in your assessments, then even if you cannot explain how you reach your ultimate conclusions, and even if therefore your process fundamentally is not scientific, you can pass a superficial reading of the *Daubert* factors.

In light of this superficial match on the *Daubert* factors, it might seem reasonable just to err on the side of admissibility. After all, there have been some suggestions that *Daubert* is being improperly used as a gatekeeper to evaluate the weight of the evidence instead of its bare admissibility. *See, e.g.,* Richard D. Friedman, *"E" is for Eclectic: Multiple Perspectives on Evidence*, 87 VA. L. REV. 2029, 2052 (2002) (". . . the law of admissibility has been called on to carry weight that should be borne—if at all—by the law of sufficiency.") (excerpted in David A. Sklansky, *Evidence: Cases, Commentary, and Problems* 470 (2003); A. Leah Vickers, *Daubert, Critique, and Interpretation: What Empirical Studies Tell Us About the Application of Daubert*,

40 UNIV. OF S.F. L. REV. 109 (2005) (cataloging scholarly critiques of *Daubert*). One could argue that the better course in such cases is to let this in and let the defense attorney take his best shots at the weaknesses. But there are, I think, several reasons why a closer analysis is warranted.

First, when there are alternative paths to the admissibility of evidence, it is often of real importance which path is used. The path for admitting opinions grounded in experience and training is *Kumho Tire*, not *Daubert*. And that matters because scientific expert testimony carries special weight with a jury, as the Government articulated when it spoke of the "CSI effect." Fair or not, a technical, non-scientific expert simply does not carry quite the same authority. Courts must be careful to whom we lend that authority.

Second, the flaws in this kind of quasi-scientific methodology are more fundamental than just the enigmatic process by which the ultimate conclusion is reached, and those flaws end up being carried over into every other factor. For example, it is difficult to accurately measure the error rate for a methodology that cannot be described in a way that renders it repeatable, and an inability to repeat and test a methodology makes it challenging to subject to peer review.

Finally, one often-overlooked value of *Daubert* is that it preserves the ability to cross-examine. If someone cloaked in the powerful robe of science tells a jury he knows something, and the reason why amounts to "trust me, I'm a scientist," then the government gains most of the power of scientific evidence at none of the cost. It becomes very difficult to engage such a witness in meaningful cross examination. But an expert who has to explain himself and his methods becomes accountable to the defense and to the jury, and meaningful cross-examination becomes possible.

In light of this backdrop, I will proceed to apply *Daubert* to Mr. Gover's testimony. On each factor, when one looks more closely at his testimony and considers the underlying purpose

of the inquiry, the methodology used in this case has problems. I will consider each factor in turn.

A. Testability

The question whether a methodology is "testable" comprises two parts: falsifiability and replicability. The former asks whether we are dealing with a proposition that can be subjected to some sort of observation or test that renders the proposition false. The latter asks whether someone else could repeat the exact methodology the expert used. As discussed below, Mr. Gover's testimony presents issues primarily with replicability.

i. Falsifiability

The idea behind falsifiability is that the methodology at issue—here, toolmark comparisons—involves a proposition that observation or testing can prove false. The proposition, for example, that aliens once visited the earth cannot be refuted as false by any test or observation. But a proposition that says there are no black swans is falsifiable by the observation of a single black swan. The question here is whether the hypothesis that Mr. Gover advances fits into the former category or the latter.

I find that the proposition that a particular shell casing was fired from a particular gun is a falsifiable hypothesis. It is possible to test whether that hypothesis is false by observing characteristics so different that a match is excluded. Whether Mr. Gover has succeeded at arriving at a scientifically reliable answer is a different question, but such an answer is possible, and I find that the stated proposition is falsifiable.

ii. Replicability

I do not, however, find that the AFTE comparison testing methodology, as described by Mr. Gover, is replicable. In order to be replicable, a method must be objective enough that

someone else not associated with the case could duplicate it and get the same results. That standard is very problematic here. Much of what Mr. Gover did was, in fact, objective. He and his colleagues observed a careful chain of custody. They took care not to contaminate the comparators—no stray marks were added in the lab by accident. They carefully calibrated the equipment to achieve accurate, side-by-side images. Mr. Gover used a powerful microscope whose functionality has been scientifically verified. He correctly verified that the shell in question matches, in caliber and otherwise, the make and model of the Taurus and could have been fired from it. He verified that the Taurus was an operable firearm that could have fired the shell casing. But when it comes to his ultimate conclusion of a match, Mr. Gover's inquiry was not objective.

The first problem is with the baseline comparator that Mr. Gover used—the "best known non-match" as a comparator. This requires an examiner to compare the degree of correspondence on his split screen against the best possible non-match comparison that anyone is aware of between consecutively manufactured firearms. Tr. [70] at 24. For example, if the best possible comparison between a non-matching shell casing and firearm would show five matching striae, Mr. Gover would be looking for at least six. But Mr. Gover could not define this baseline in any objective way, nor could he explain the role it played in the actual comparison he made in this case. He apparently just kept a memory of the baseline in his head and then made a comparison unguided by any objective standards or benchmarks. For example:

> [THE COURT]: So what characteristics make two shells a better match than the best known non-matching shells?
>
> [MR. GOVER]: What characteristics?
>
> [THE COURT]: Yes.

> [MR. GOVER]: That's part of the consecutively manufactured studies is being able to look at, say, the piece that made these marks on one gun and then looking at one that was manufactured with the same tool right afterwards. That's where we expect the most agreement between two different, because it's the same tool, one right after the other that's producing the two unknowns.
>
> That's where we expect to see the most agreement between two different tools, two different aspects of the firearm. That's where all these consecutive manufactured studies play their part. When you start looking at all of these comparisons, you'll see very limited corresponding detail between different firearms, and then you'll see correspondence between two dates from the same firearm, which is what in my opinion this exemplifies.
>
> [THE COURT]: So you have in your head a benchmark above which your two shells have to exceed –
> [MR. GOVER]: Yes.
>
> [THE COURT]: – this to be a match. You have that available to you. That's really the core of your methodology, right?
>
> [MR. GOVER]: Yes.
>
> [THE COURT]: I guess I appreciate that as your core, I just don't know what it is.
>
> [MR. GOVER]: It's hard to express, I guess. That's where I think that consecutively matching striae criteria, it does have a place to help supplement because it's been tested numerically in the past.

Tr. [70] at 60-61.

It is impossible to glean any meaningful benchmark from this exchange. Nor could Mr. Gover express any objective benchmark at any point during his testimony. With no objective benchmark below which we would know that a shell casing was not fired from a particular firearm (other than the CMS methodology that he elsewhere expressly disavowed as the basis for his opinion), replicability begins to fall apart.

Throughout Mr. Gover's testimony, this lack of any objective standard became increasingly evident. It is worth explaining at the outset that this flaw in Mr. Gover's testimony appears to derive from the AFTE methodology itself, not from any deficiency in Mr. Gover's

understanding of his work. AFTE requires an examiner to find "sufficient agreement" between crime scene shells and test fired shells from the firearm in question in order to determine a match. Tr. [70] at 23. Here is Judge Garaufis of the Eastern District of New York explaining why "sufficient agreement" is not an objective standard:

> First, the sufficient agreement standard is circular and subjective. Reduced to its simplest terms, the AFTE Theory "declares that an examiner may state that two toolmarks have a 'common origin' when their features are in 'sufficient agreement.'" *PCAST Report* at 60. "It then defines 'sufficient agreement' as occurring when the examiner considers it a 'practical impossibility' that the toolmarks have different origins." Id. The NRC Report notes that the AFTE Theory "is the best guidance available for the field of toolmark identification, [but] does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence." *NRC Report* at 155. Without guidance as to the extent of commonality necessary to find "sufficient agreement," the AFTE Theory instructs examiners to draw identification conclusions from what is essentially a hunch—a hunch "based on the examiner's training and experience," *AFTE Revised Theory of Identification*, 43 AFTE Journal at 287—but still a hunch.
>
> Moreover, the application of this circular standard is "subjective in nature ... based on the examiner's training and experience." *AFTE Revised Theory of Identification*, 43 AFTE Journal at 287. Ostensibly, one hundred firearms toolmark examiners could hold one hundred different personal standards of when two sets of toolmarks sufficiently agree, and all one hundred of these personal standards may accord with the AFTE Theory. Further, because the standard itself offers so little guidance on when an examiner should make an identification determination, some examiners may decide that the two sets of toolmarks were made by the same tool while others determine the toolmarks to be inconclusive and still others decide the toolmarks were made by different tools. To emphasize, these one hundred examiners could come to these contradictory conclusions without a single examiner running afoul of the AFTE Theory.

United States v. Shipp, No. 19-cr-029-NGG, 2019 WL 6329658 *13 (E.D.N.Y. Nov. 26, 2019).

In other words, the AFTE "sufficient agreement" standard is a tautology that doesn't *mean* anything. This was evident throughout Mr. Gover's testimony as he struggled to explain what he was looking for in order to conclude the shell casings matched the Taurus. For example:

[PROSECUTOR]: So if there's not a numeric threshold when you're doing this identification, what assurances do you have that your conclusions will be consistent with the conclusions of other toolmark examiners?

[MR. GOVER]: Part of it is I'm using the same methodology, the same criteria for identification, and once I've completed doing my actual comparative analysis and I have drawn a conclusion, within our system, a second qualified firearms examiner will come in right behind me and look at that same evidence to either agree or disagree with my findings.

[PROSECUTOR]: And is it fair to say that AFTE has described the sufficient agreement as being a subjective standard

[MR. GOVER]: Yes. The interpretation of the objective observations, which is viewing those contours, surface contours of the toolmarks, being able to line up the peaks and valleys and ridges and furrows, that's – those are all objective observations. The interpretation of the observation is what is the subjective aspect of the firearms identification.

[PROSECUTOR]: And when you're making your conclusion, is that based in part on your experience?

[MR. GOVER]: Based on experience, training, research provided by the Association of Firearm and Toolmark Examiners, current training, current validations that are going on now, as well as past experiments which date back decades, you know, being done by AFTE. So it's a combination of all of it.

Tr. [70] at 25-26.

[PROSECUTOR]: And was it your conclusion that there was sufficient agreement between those two?

[MR. GOVER]: Yes.

[PROSECUTOR]: And just for our benefit, can you sort of show your math, show us how you got there, in terms of just when you're looking at those two photos, what is it in particular that you're looking for that leads you to conclude, hey, there's sufficient agreement here?

[MR. GOVER]: I'm looking at the correspondence of those striated marks between my test fire on the left and my unknown cartridge case on the right and evaluating that correspondence. And from what I can see from that correspondence, that exceeded the known level of individual characteristics that I would expect to see or correspond between two cartridge cases fired in two different firearms.

. . .

[PROSECUTOR]: I just want to go back over this one more time. I think we talked about this before. But why not, when doing your analysis to see if there's sufficient agreement, why not have a numeric threshold that if there's, say, seven striations that match, that's enough, but below that isn't good enough?

[MR. GOVER]: Why not?

[PROSECUTOR]: Yeah. Why isn't there a numeric – why can't we qualify this with numbers?

[MR. GOVER]: Part of it for me, to understand that quantifiable is almost another subjective aspect of when I'm looking at these furrows and ridges, what constitutes the consecutively – what they refer to as the consecutively manufactured or consecutively matching striae. So I don't have a complete answer to that and I've tried to formulate an argument as to why or why not, and I haven't, you know, heard a lot of other people talk about it."

Tr. [70] at 38-40.

[DEFENSE ATTORNEY]: . . . . And I'm sorry, the name of the person that did the review after you and confirmed?

[MR. GOVER]: Allesio.

[DEFENSE ATTORNEY]: Allesio. Now, the threshold that – so is it fair to say that Mr. Allesio has his own threshold in terms of what is sufficient agreement?

[MR. GOVER]: I don't – if he does, I would assume that it's pretty much the same, being as we have the same level of experience and type of training.

[DEFENSE ATTORNEY]: But you have no way of knowing that?

[MR. GOVER]: No.

[DEFENSE ATTORNEY]: Because it's based on his own perceptions and what he happened to retain?

[MR. GOVER]: It depends on his interpretation of the objective –

[DEFENSE ATTORNEY]: Maybe he read some literature that you didn't read, that kind of thing?

[MR. GOVER]: We both receive the AFTE Journal, so it's available for both of us.

Tr. [70] at 46-47.

This last excerpt of testimony is particularly damaging to the admissibility of this methodology under *Daubert*. Mr. Gover could not say that the person who checked his work, and who relied on the same methodology, applied the same standard in reaching the same conclusion. He could not be sure what threshold Mr. Allesio used to decide that the shell casings were fired from the Taurus—*i.e.* that Mr. Gover's conclusion was correct. Not only is the AFTE method not replicable for an outsider to the method, but it is not replicable between trained members of AFTE who are using the same means of testing.

If this were truly a scientific inquiry, such testimony would not be possible. If a cancer researcher sought a second opinion from another cancer researcher in order to reach a diagnosis, both people would be able to say with certainty what the other person was looking for and why. If their conclusions deviated, they would be able to pinpoint the points of disagreement and why those data points were meaningful.

Over and over, Mr. Gover failed to do this. He could not explain which data points he looked at or why they were meaningful to him.[6] And this is not purely a fault of Mr. Gover. There is no evidence in this record or elsewhere that the AFTE method relies on any scientific standard that would explain to an examiner like Mr. Gover how to interpret the data he sees in any kind of objective way. What he is actually doing is applying his training and experience to make a subjective conclusion about what he sees before him, just like the art expert in Malcolm Gladwell's example. The AFTE method is therefore not replicable—and not testable—because it cannot be explained in a way that would allow an uninitiated person to perform the same test in the same way that Mr. Gover did. This factor weighs heavily against admissibility under *Daubert*.

---

[6] The full transcript of Mr. Gover's testimony is attached to this opinion as Appendix A.

B. Error Rates

The next *Daubert* factor that must be applied is the rate of error for the methodology in question. It is important to note that whether a particular error rate is high or low has no independent meaning outside of its context. Whether an error rate is "high" or "low" depends on the use to which the testimony will be put. Here, the use of forensic toolmark testing is very often the most critical factor in determining guilt or innocence—an endeavor that must tolerate only a very small rate of error. Even an error rate of five percent, which would be low in many contexts, would be unacceptably high where it drives a guilty verdict because it would mean that one in twenty convictions could be wrong.[7] The question here is whether the error rate for toolmark testing is acceptably low so as to be a reliable means of determining guilt or innocence.

The Government initially asserted that the error rate for toolmark comparison testing is between .9 and 1.5 percent. Gov't. Resp. [57] at 9. But testing shows a range of outcomes, sometimes with an error rate as high as 2.2 percent. *United States v. Shipp*, No. 19-cr-029-NGG, 2019 WL 6329658 *12 (E.D.N.Y., Nov. 26, 2019). If these all sound like low rates of error, whose differences could not possibly be material, it is helpful to consider them in terms of wrongful convictions, which is the correct framework for an error rate that measures only false-positives—*i.e.* incorrectly identified matches. *See, e.g.*, President's Council of Advisors on Sci. & Tech., *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature*

---

[7] I recognize that the "one in twenty" figure would be an overstatement if the error rate included both false positives and false negatives—*i.e.* resulting in false acquittals as well as false convictions. But as I explain below, I think the testing methods for the error rate of the AFTE method yield results almost exclusively focused on false positives. See, e.g., *PCAST Report*, 104-114 (discussing firearms testing). The PCAST Report was a "meta-study," which means it examined the results of all the fields tests that had been done and drew conclusions from all of those studies in context of each other. All of the studies cited in the PCAST Report focused on false positives when reaching an error rate for ballistics comparison testing.

Comparison Methods, 104-14 (2016) ("PCAST Report") (discussing firearms testing). A .9 percent error rate would lead to about 1 in 111 wrongful convictions. A 1.5 percent error rate would mean that 1 in 67 convictions were wrong. And 2.2 percent would mean that 1 in 46 convictions were wrong. These are dramatically different rates of error when put into context.

What's more, the higher error rates tend to arise from the studies that most closely resemble the real-world conditions of toolmark testing. The lowest rates arise from the "closed-set" tests, which require the examinee to perform a matching exercise between two sets of bullets or shell casings. *Shipp*, 2019 WL 6329658 at *12 (citing *PCAST Report*, 106-11 (2016)). An examinee can "perform perfectly" if he simply matches each bullet to the standard that is closest. *Id.* Further, each match narrows the field for further matches. *Id.*

The next highest error rates—about 2.1 percent—arise from partly closed sets. *Id.* (citing PCAST Report at 109). These tests also give the examinee a closed set of matches, but it also includes two bullets or shells that do not have a match in the set. *Id.* The error rate from these tests is "nearly 100-fold higher" than from the closed-set tests. *Id.*

Finally, the "black box" studies yield the highest error rates, about 2.2. percent. *Id.* (citing PCAST Report at 110-11). These tests presented each examinee with an unknown shell casing or bullet and three test fires from the same known firearm, which may or may not have been the source of the unknown casing or bullet. *Id.* These tests most closely resemble real-world analysis—*i.e.* what Mr. Gover testified that he did in this case.

On the other hand, Mr. Gover testified that a study from the Ames Laboratory at Iowa State University found a 1 percent error rate from a test in which 218 examinees were given cartridges fired from 25 different firearms and used the methodology at issue in this case to match the casings to the firearms. Tr. [70] at 27. The details of this study are not clear from Mr.

Gover's testimony, but this appears to have been a "closed-set" test, the type with the lowest error rates on average.

The incentive structure for the testing process is also concerning. It appears to be the case that the only way to do poorly on a test of the AFTE method is to record a false positive. There seems to be no real negative consequence for reaching an answer of inconclusive.[8] Since the test takers know this, and know they are being tested, it at least incentivizes a rate of false positives that is lower than real world results. This may mean the error rate is lower from testing than in real world examinations.

It is hard to know exactly what to make of these results. It is possible that the error rate for toolmark testing is very low, but it is more likely that it is not. Assuming false positive test results lead to wrongful convictions, a wrongful conviction rate of 1 in 46 is far too high. The best test results would favor the government, but it is unlikely those tests reflect real-world error rates. The worst results favor Defendant. At most, then, this factor of the *Daubert* test is neutral as to both parties. In my opinion, it cuts somewhat in favor of Defendant.

C. Peer Review

At the outset, it is important to remember the purpose of this *Daubert* factor. The question of whether a methodology has been subjected to peer review is a question of whether a methodology or hypothesis has been published to the scientific community for the purpose of detecting substantive flaws. *Daubert*, 509 U.S. at 593. If a methodology has been published but not for this purpose, this factor is not satisfied in favor of admissibility.

---

[8] In fact, the closed-set study discussed above yielded an "inconclusive" rate of 41.8 percent, and the black box study yielded an inconclusive rate of 33.7 percent. *PCAST Report* at 111. These results were not included in the "error rate."

Here, Mr. Gover testified that the methodology he uses is peer reviewed through the AFTE Journal, which exercises quality control over studies that are done and decides whether they are "worth publication." Tr. [70] at 31. The Journal does not evaluate the methodology itself, which has been established as the industry standard since 1992. *Id.* at 32. The only question the Journal asks is whether any studies being published used the correct, accepted methodology. *Id.* This does not amount to peer review, for two reasons. First, the AFTE Journal is a trade publication, meant only for industry insiders, not the scientific community. Second and more importantly, the purpose of publication in the AFTE Journal is not to review the methodology for flaws but to review studies for their adherence to the methodology. In fact, the methodology has never changed, aside from a minor revision of terminology, in the 18 years Mr. Gover has worked as a forensic scientist. *Id.* at 32-33. This is not the purpose that *Daubert* sets out for peer review. This factor therefore favors Defendant.

D. Standards and Quality Control

Mr. Gover did identify some quality control mechanisms. For example, he takes an annual proficiency test. Tr. [70] at 33. The test yields only binary pass/fail results, not a rate of error, but it does give Mr. Gover some information about how well he can identify firearms matches. *Id.* Further, every forensic toolmark test is reviewed by a second examiner who either verifies or disagrees with the conclusion. *Id.* Then the results are put through a "technical review" process, in which another examiner reviews the notes taken during the comparison testing phase in order to make sure that the proper procedures were followed and that the examiners' conclusions are supported. *Id.* at 34. At a more general level, Mr. Gover and his colleagues also receive training and procedures manuals from AFTE that explain how to do comparison testing and what processes to follow. *Id.*

This amounts to quality control. If these enforcement mechanisms were applied to a scientifically testable methodology, it would be easy to say that toolmark comparison testing was held to a high standard and subject to quality control. This *Daubert* factor therefore favors the Government, although it is not dispositive.

E. General Acceptance

The fifth and final *Daubert* factor asks whether the methodology in question has been accepted in the broader scientific community. This is a difficult question to answer. The AFTE method that Mr. Gover uses has been widely accepted within his own community of technical experts. Tr. [70] at 35. But it has been heavily criticized by other members of the broader scientific community for failing to yield reproducible results or a precisely defined process. Def.'s Br. [55] at 6-7 (citing Nat'l Research Council, *Ballistic Imaging* 81 (National Academies Press 2008)); NRC Comm. on Identifying the Needs of the Forensic Sci. Cmty., *Strengthening Forensic Science in the United States: A Path Forward* 155 (2009)). In fact, these reports suggest to me that the widespread acceptance within the law enforcement community may have created a feedback loop that has inhibited the AFTE method from being further developed. On the other hand, widespread general acceptance within a given technical community could theoretically be sufficient. But that is more consistent with a *Kumho Tire* theory of expert testimony, not with *Daubert*. Here, where the scientific community at large disavows the theory because it does not meet the parameters of science, I cannot find that the AFTE method enjoys "general acceptance" in the scientific community.

CONCLUSION

I want to be clear that my ruling, as expressed in the foregoing opinion, is limited by the testimony before me during the hearings held in this case. It is not an indictment of forensic

evidence or toolmark comparison analysis writ large. It is clear that Mr. Gover and his colleagues are on to something. Even at its worst, comparison analysis has a very low rate of error and yields results that cannot be random. But it is not clear that those results are the product of a *scientific* inquiry. Nothing in Mr. Gover's testimony explains how or why he reached his conclusion in any quantifiable, replicable way. It is possible that the AFTE method could be expressed in scientific terms, but I have not seen it done in this case, nor elsewhere.[9]

Therefore, for the reasons discussed above, Mr. Adams's Motion *in Limine* [55] is GRANTED in part and DENIED in part. Mr. Gover's expert testimony is limited to the following observational evidence: (1) the Taurus pistol recovered in the crawlspace of Mr. Adams's home is a 40 caliber, semi-automatic pistol with a hemispheric-tipped firing pin, barrel with six lands/grooves and right twist; (2) that the casings test fired from the Taurus showed 40 caliber, hemispheric firing pin impression; (3) the casings seized from outside the shooting scene were 40 caliber, with hemispheric firing pin impressions; and (4) the bullet recovered from gold Oldsmobile at the scene of the shooting were 40/10mm caliber, with six lands/groves and a right twist.

//

//

---

[9] After listening at length to Mr. Gover's testimony, it appears there is an element of over-inclusiveness at play. It seems likely that the AFTE method could be revised to rest on quantitative factors. It seems equally likely that a more quantitative measure of sufficient agreement would result in a finding of inconclusive in cases that currently result in a match. Mr. Gover and his peers seem reluctant to impose quantitative restrictions on their methodology because it would fail to justify a match in those cases where the numerical standard isn't met, but the trained examiner has an impression—call it a hunch—that it is actually a match. But there are several settings in which law enforcement officials are required to leave their hunches at the courthouse door. It is only slightly inaccurate to say that the criminal justice system is designed to favor false negatives over false positives. To be admissible, in such a system, as scientific evidence, AFTE will have to shift away from hunches to numbers.

No evidence relating to Mr. Gover's methodology or conclusions relating to whether the shell casings matched the Taurus will be admitted at trial.

IT IS SO ORDERED.

DATED this 16th day of March, 2020.

MICHAEL W. MOSMAN
United States District Judge